Generative Face Completion

Yijun Li, Sifei Liu, Jimei Yang, and Ming-Hsuan Yang

STRUCT Paper Reading 杜昆泰

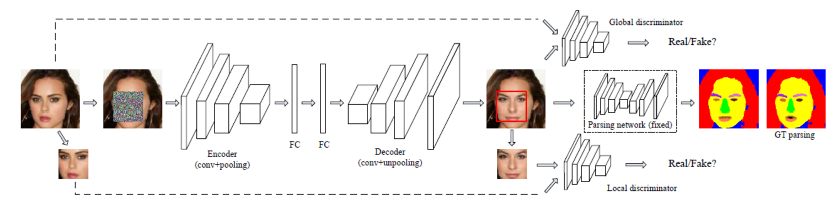
2017 年 10 月 20 日

Face completion is a difficult task.

- It requires to generate semantically new pixels for the missing key components.
- Many object parts in the input image contain unique patterns.

GAN of course = =

The architecture is shown as below:



Generator:

”conv1” to ”pool3” in vgg19

+ 2 conv

+ 1 pooling

+ 1 fc

+ symmetric decoder

Discriminator:

Local: see if the reconstructed part seems real

Global: see if the whole reconstructed image seems real

Semantic regularization: main contribution

Q: how to ensure the consistency in the generated image? (e.g.: the real eye and the generated eye must be alike)

A: Construct a loss to ensure that the semantic parsing result of the whole generated image is alike similar to the result of GT.

Loss function

$$L = L_r + \lambda_1 L_{a_1} + \lambda_2 L_{a_2} + \lambda_3 L_p$$

Where L_r is simply the L_2 loss between the generated result and GT, L_{a_1} is the local GAN loss, L_{a_2} is the global GAN loss, L_p is the softmax loss between the parsing result of generated image and GT.

Softmax loss: Use $-log(\frac{e^y}{\sum_{j=1}^{m} e^j})$ to maximize the softmax probability of class y.

Training Method:

- step1. simply by L_2 loss
- step2. L_2 loss + local adversarial loss
- step3. All loss

On CelebA test dataset:

Result of different settings:

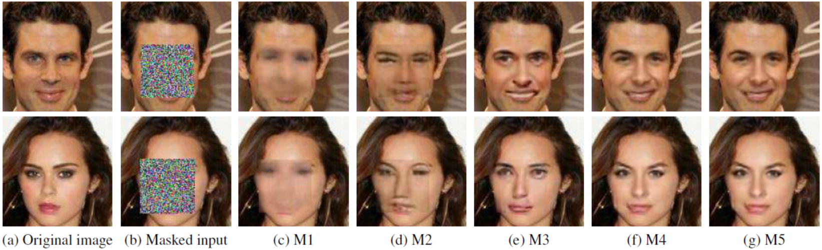


Figure 3. Completion results under different settings of our model. (c) M1: L_r. (d) M2: $L_r + L_{a_1}$. (e) M3: $L_r + L_{a_1} + L_{a_2}$. (f) M4: $L_r + L_{a_1} + L_{a_2} + L_p$. The result in (f) shows the most realistic and plausible completed content. It can be further improved through post-processing techniques such as (g) M5: M4 + Poisson blending [18] to eliminate subtle color difference along mask boundaries.

Semantic parsing

(a) original (b) masked input (c) w/o parsing (d) w/ parsing

Figure 4. Comparison between the result of models without and with the parsing regularization.

Quantitative result:

Table 1. Quantitative evaluations in terms of SSIM at six different masks O1-O6. Higher values are better.

	M1	M2	M3	M4	CE	M5
O1	0.798	0.753	0.782	0.804	0.772	**0.824**
O2	0.805	0.763	0.787	0.808	0.774	**0.826**
O3	0.723	0.675	0.708	0.731	0.719	**0.759**
O4	0.747	0.701	0.741	0.759	0.754	**0.789**
O5	0.751	0.706	0.732	0.755	0.757	**0.784**
O6	0.807	0.764	0.808	0.824	0.818	**0.841**

Table 2. Quantitative evaluations in terms of PSNR at six different masks O1-O6. Higher values are better.

	M1	M2	M3	M4	CE	M5
O1	18.9	17.8	18.9	19.4	18.6	**20.0**
O2	18.7	17.9	18.7	19.3	18.4	**19.8**
O3	17.9	17.2	17.7	18.3	17.9	**18.8**
O4	18.6	17.7	18.5	19.1	19.0	**19.7**
O5	18.7	17.6	18.4	18.9	19.1	**19.5**
O6	18.8	17.3	19.0	19.7	19.3	**20.2**

Table 3. Quantitative evaluations in terms of identity distance at six different masks O1-O6. Lower values are better.

	M1	M2	M3	M4	CE	M5
O1	0.763	0.775	0.694	0.602	0.701	**0.534**
O2	1.05	1.02	0.894	0.838	0.908	**0.752**
O3	0.781	0.693	0.674	0.571	0.561	**0.549**
O4	0.310	0.307	0.265	0.238	0.236	**0.212**
O5	0.344	0.321	0.297	0.256	0.251	**0.231**
O6	0.732	0.714	0.593	0.576	0.585	**0.541**

The model still cannot generate plausible result with unaligned face.

The spacial correlation of adjacent pixels does not fully exploited by the model.

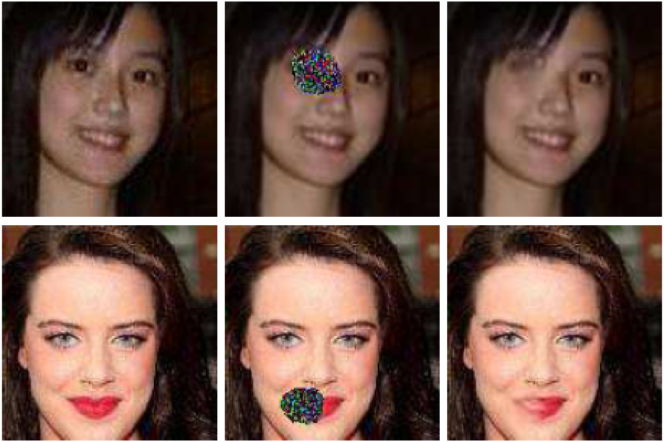


Figure 12. Model limitations. Top: our model fails to generate the